DOCUMENT RESUME

ED 459 836 IR 058 372

AUTHOR	Ben-Arie, Jezekiel; Pandit, Purvin; Rajaram, ShyamSundar
TITLE	Design of a Digital Library for Human Movement.
SPONS AGENCY	National Science Foundation, Arlington, VA.
PUB DATE	2001-06-00
NOTE	12p.; In: Proceedings of the ACM/IEEE-CS Joint Conference on Digital Libraries (1st, Roanoke, Virginia, June 24-28, 2001). For entire proceedings, see IR 058 348. Figures may not reproduce well.
CONTRACT	IIS-9711925; IIS-9876904; IIS-9979774
AVAILABLE FROM	Association for Computing Machinery, 1515 Broadway, New York NY 10036. Tel: 1-800-342-6626 (U.S. & Canada); Tel: +1-212-626-0500 (Global); e-mail: acmhelp@acm.org. For full text: http://www1.acm.org/pubs/contents/proceedings/dl/379437/.
PUB TYPE	Numerical/Quantitative Data (110) -- Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Computer System Design; Database Design; *Electronic Libraries; *Human Posture; Information Retrieval; Information Systems; Library Development; Video Equipment; Visual Aids
IDENTIFIERS	*Video Technology

ABSTRACT

This paper is focused on a central aspect in the design of a planned digital library for human movement, i.e. on the aspect of representation and recognition of human activity from video data. The method of representation is important since it has a major impact on the design of all the other building blocks of the system such as the user interface/query block or the activity recognition/storage block. This paper evaluates a representation method for human movement that is based on sequences of angular poses and angular velocities of the human skeletal joints, for storage and retrieval of human actions in video databases. The choice of a representation method plays an important role in the database structure, search methods, storage efficiency, et cetera. For this representation, the study develops a novel approach for complex human activity recognition by employing multi-dimensional indexing combined with temporal or sequential correlation. This scheme is then evaluated with respect to its efficiency in storage and retrieval. For the indexing, the study uses postures of humans in videos that are decomposed into a set of multidimensional tuples which are present the poses/velocities of human body parts, such as arms, legs and torso. Three novel methods for human activity recognition are theoretically and experimentally compared. The methods require only a few sparsely sampled human postures. Speed invariant recognition of activities is also achieved by eliminating the time factor and replacing it with sequence information. The indexing approach also provides robust recognition and an efficient storage/retrieval of all the activities in a small set of hash tables. (Contains 24 references.) (Author/AEF)

Design of a Digital Library for Human Movement

By: Jezekiel Ben-Arie, Purvin Pandit & Shy amSundar Rajaram

Design of A Digital Library for Human Movement *

Jezekiel Ben-Arie
EECS Department M/C 154
University of Illinois at Chicago
851 S. Morgan St., SEO 1120
Chicago, IL 60607, USA.
benarie@eecs.uic.edu

Purvin Pandit
EECS Department
University of Illinois at Chicago
ppandit@eecs.uic.edu

ShyamSundar Rajaram
EECS Department
University of Illinois at Chicago
srajaram@eecs.uic.edu

ABSTRACT

This paper is focused on a central aspect in the design of our planned digital library for human movement, i.e. on the aspect of representation and recognition of human activity from video data. The method of representation is important since it has a major impact on the design of all the other building blocks of our system such as the user interface/query block or the activity recognition/storage block. In this paper we evaluate a representation method for human movement that is based on sequences of angular poses and angular velocities of the human skeletal joints, for storage and retrieval of human actions in video databases. The choice of a representation method plays an important role in the database structure, search methods, storage efficiency etc.. For this representation, we develop a novel approach for complex human activity recognition by employing multidimensional indexing combined with temporal or sequential correlation. This scheme is then evaluated with respect to its efficiency in storage and retrieval.

For the indexing we use postures of humans in videos that are decomposed into a set of multidimensional tuples which represent the poses/velocities of human body parts such as arms, legs and torso. Three novel methods for human activity recognition are theoretically and experimentally compared. The methods require only a few sparsely sampled human postures. We also achieve speed invariant recognition of activities by eliminating the time factor and replacing it with sequence information. The indexing approach also provides robust recognition and an efficient storage/retrieval of all the activities in a small set of hash tables.

Categories and Subject Descriptors

I.4.8 **[Image Processing and Computer Vision]**: Scene Analysis—*Motion, Tracking*; I.5.2 **[Pattern Recognition]**: Design Methodology—*Pattern Analysis*; E.2 **[Data Storage Representation]**: Hash-table representation

General Terms

Design,Algorithms

Keywords

Human Activity Recognition, Multi Dimensional Indexing, Temporal correlation, Sequence Recognition

1. INTRODUCTION

Human movement analysis is an important ingredient in many areas such as kinesiology, biomechanics, rehabilitative procedures, ergonomic evaluations of job tasks, anthropology, cultural studies, sign language, athletic analysis and sports medicine. In addition, research in artistic areas such as dancing, gymnastics, figure skating, ethnic studies and behavioral studies also require analysis, representation and classification of human motion. There are many libraries that include video and other motion data (like dance choreography notations [7]) that can be utilized as sources of raw data. However, such raw data do not accurately quantify and represent the actual three dimensional human motion, which can be quite complex. Furthermore, there is no universal method for accurate and detailed representation of human motion that can be employed to uniquely define search queries for generic or particular types of human actions/activities. Obviously, there is a need for a general method for representation of human action that can be employed to uniquely specify, store and retrieve such actions in digital libraries.

The overall architecture of our planned digital library which is called as HUman MOtion Retrieval System (HUMOR-S) is shown in Fig. 1. In this figure the system is divided into three parts, the user interface, the motion/action recognition module and the motion sequence and learning module. This paper is focused on the motion/action recognition module which enables to recognize human activity from video sequences using hash table representation. The other two modules are described briefly to outline the complete structure of the system. The user interface module handles the user's queries and provides a graphic feedback of humanoid animation for interactive querying.This module is designed to accept three modes for query input: spatio-temporal, symbolic and visual. In the symbolic query mode, the user can specify a query which is composed of a sequence

*This work was supported by the National Science Foundation (NSF) Grants No. IIS-9711925, IIS-9876904 and IIS-9979774.

JCDL'01, June 24-28, 2001, Roanoke, Virginia, USA.
Copyright 2001 ACM 1-58113-345-6/01/0006 ...$5.00.

of basic actions from a menu. A basic action could be raising a hand or nodding the head. Such basic actions are similar to the actions that are defined in dance notations such as Labatonian Sutton Dance notation [7]. This input mode is quite limited and may not include all the possible human actions. Whenever necessary, it is proposed to add new actions to the actions menu by composing them with the spatio-temporal query module.

The user interface module also has a temporal sequencer which can combine these basic actions into more complete human activities. These sequences can be fed into a humanoid animation module which displays the resulting motion sequence and provides visual feedback to the user. The second mode of input query is the spatio-temporal mode. This mode is required for articulated and accurate specification of human motion. In this mode, the user quantitatively specifies angles and velocities of specific body parts which take part in the specified action. The third mode for motion query is query by visual example, where the query is selected from a video database or is presented from an external video source. One option for an external video source might be a video camera which captures the user himself who can directly demonstrate what he wants. The third module is used for storage and learning. This is a database of all the actions/activities known to the system. A new sequence may be added to the database, whenever it is sufficiently dissimilar to any of the existing sequences in the database. By this manner, the system is capable of learning new motion sequences.

The objective of this paper is to perform an initial assessment of the feasibility of a proposed representation method for human movement that is based on angular poses and angular velocities of the human skeletal joints, for analysis, querying and retrieval of human actions in video databases that describe human motion. The representation method plays a major role in determining the overall database structure, search methods, storage efficiency and other important facets and therefore any choice has to be evaluated very carefully.

Natural languages and symbolic temporal descriptions are not suited to **accurately** describe **articulated** human movements or actions. Natural language enables to describe human actions only in generic terms such as "walking", "running", "swimming", etc. Such actions constitute in reality, a sequence of complex motions of body parts that may differ from person to person. Other languages based on symbolic representations such as Spatio-Temporal Logic (STL) [5], Hierarchical Temporal Logic (HTL) [22], or Symbolic Projection based languages [3] [12] are also quite limited and can describe only gross spatial relations and motions between different rigid objects. Obviously, the human body is not a rigid object and its actions cannot be specified in such a limited representation. There are other notations that were developed to describe human motion in actions such as dancing [7] [8], figure-skating[9], athletic exercises[9] and natural gestures[24]. However, all of these notations are useful only for their specific domain and cannot universally describe articulated human motion. In addition, these methods also are based on symbolic representations and therefore quantize the motion of human body parts very coarsely.

Recently, the Virtual Reality Modeling Language (VRML) is being developed as a tool for animation of humanoids in networked applications (http://ece.uwaterloo.ca/~h-anim/) in conjunction with the development of MPEG-4 / SNHC [1] (http://www.es.com/mpeg4-snhc/)

The VRML humanoid provides a framework for accurate representation of human motion by enabling to define the angular poses and velocities of all the joints in the human body. In this work, we learn from the general structure of the VRML humanoid model and we propose a vectorial representation of human motion that is capable of accurate and detailed description by a sequence of multi-dimensional vectors. This proposed representation is elaborated in Section 2. As detailed in Section 4, we develop an efficient methods to store and index such vectorial sequences. This enables accurate analysis of articulated human activity/action as well as fast retrieval and articulated formulation of queries.

In order to evaluate the efficacy of our proposed angular pose/velocity representation , we develop and compare several approaches for the recognition of human activity based on indexing of sparsely sampled angular poses/velocities of the limbs and the torso. The sampled poses/velocities are obtained by tracking body parts in video sequences. We develop in this paper three indexing based methods for human activity recognition that differ in the pose and the temporal information used.

Human body's static posture frequently gives an indication of the action that takes place. This fact is evident from observing static images that reveal in many cases the actual activity. Based on this idea, we examine in this paper the possibility of recognizing human activities just from few sampled postures. A person's posture is composed of the poses of arms, legs, torso and head. Human activity can be described as a temporal sequence of pose vectors that represent sampled poses of body parts. Our principle of recognizing human activity from sparsely sampled postures is based on identifying these postures as samples of a complete, densely sampled model activity. To achieve this objective, we construct a database that includes all the model activities in the form of entries in multidimensional hash tables. The size of these tables is not too large since, body parts have limited angular motion and thus the number of bins that describe the full range of angular motion of each body part is quite limited.

An important feature of our approach is the separation of the multi-dimensional indexing into several hash tables, where each table corresponds to a different body part. This structure enables to index and recognize activities even when several body parts are occluded. Also, our approach of using multidimensional tuples proves to be very efficient in terms of storage since all the activities are stored in the same table. Experimental results described in Section 5 demonstrate robust recognition of activities.

Several approaches for activity recognition have been reported in the literature. However, none of these works aimed at a complete human activity recognition as is demonstrated in this paper. Schlenzig, Hunter and Jain [20] use Hidden Markov Model (HMM) and a rotation-invariant imaging representation to recognize visual gestures such as "hello" and "good-bye". HMMs are also utilized by Starner and Pentland [23] to recognize American Sign Languages (ASL). In

[1] The MPEG-4 standard is focused on content based coding of video objects such as animated humanoids whereas the new MPEG-7 standard which is designed for the far future is expected to address the issues of content based indexing and retrieval.

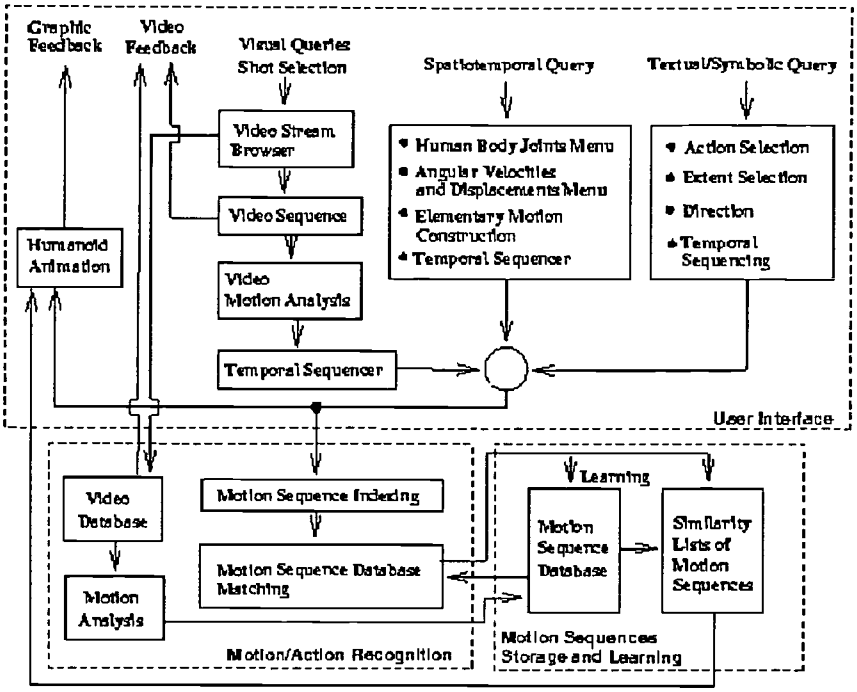

Figure 1: Architecture for the HUman MOtion Retrieval System (HUMOR-S)

these approaches, both the model and the test activities need to be densely sampled over time in order to maintain high recognition rate.

In this paper, we present a novel approach to complex human activity recognition by employing multidimensional indexing combined with temporal or sequential correlation. The representations of human activity models are actually sequences of body part poses. To be more specific, the postures of humans in each video frame are decomposed into a set of multidimensional tuples which represent the poses of human body parts such as arms, legs and torso. Whenever possible, the poses of the body parts are converted into a set of normalized angles to achieve size invariance. Hence, human activity is represented by a temporal or sequential arrangement of sets of multidimensional tuples that correspond to sampled angular poses of body parts over the entire time interval. The final outcome of this arrangement transforms human activity information into a set of hash tables each of which corresponds to an individual body part. The indices in these hash tables are the poses of the corresponding parts and the contents of those hash tables are the identities of the model activities and their time labels. At the recognition stage, a set of multidimensional indices are derived from the video sequence of the test activity.

As elaborated in the following sections, each video frame in the test activity sequence yields a 1D vote vector for each activity model in the database. The overall vote for each activity model is obtained by integrating the votes for all the test frames using temporal or sequential correlation. Details are provided in Section 4. One of the main advantages of this approach is the tremendous flexibility it provides in sampling the test activity sequence. There are no strict requirements either on the number of frames sampled or on the frame intervals of the test sequence. Users need only a set of sparsely sampled representative frames for activity recognition. This is especially useful for human activity retrieval from a large database. Our organization of the activity database also results in tremendous reduction of space requirement and significant simplification over other activity representations that usually require an explicit representation of all the activities. Experimental comparison of the three methods shows that the sequential method augmented with velocity data achieves the best results.

In Section 2 we introduce our proposed representation for human activity/action. The theoretical foundations of our approach is discussed in Section 3. In Section 4 we examine this representation in the major requirements of such a digital library, i.e. efficient storage and retrieval. The assessment is performed by developing a multidimensional indexing scheme for activity retrieval and storage. Section 5 discusses the application of our approach and demonstrates the method experimentally.

2. A PROPOSED SPECIFICATION OF HUMAN MOTION USING STATE VECTORS

In this section, we propose a representation of human motion that can be employed to accurately represent elementary motions which can be later composed to form complete human actions. This representation is conceptually similar to the approaches used for modeling human motion [2] [10] [18] [17] [6] [4] [11] [13]. Such representations are based on robot motions [15] [19]. The VRML humanoid specification provides us with a list of body joints and their relations

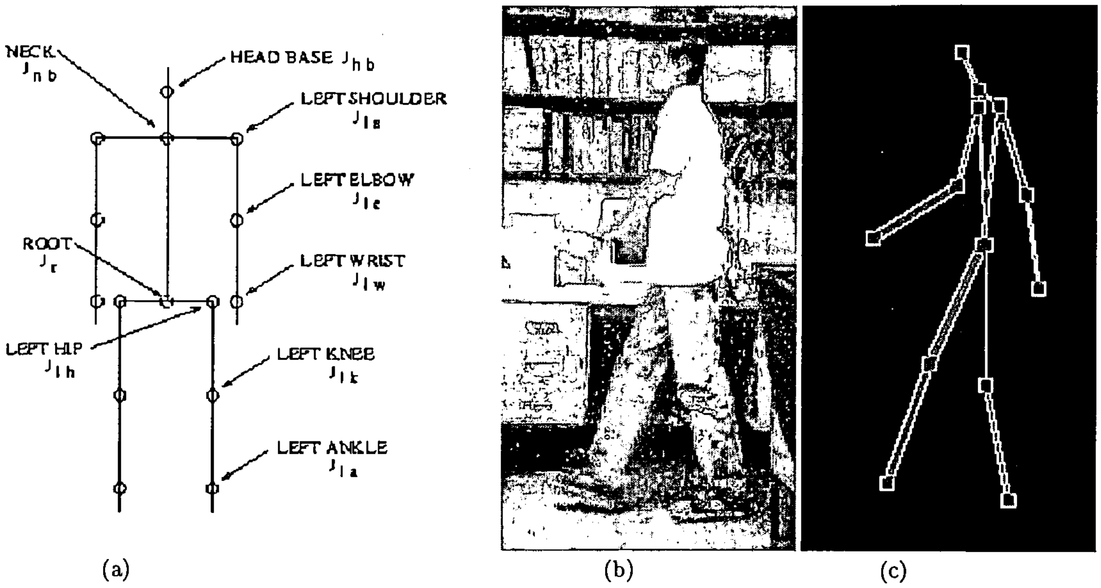

Figure 2: (a) Skeletal structure for humanoid animation.Joints are shown with the corresponding motion vector. (b) Human model in a walk posture. (c) Human Skeletal model for walk posture.

which can be expressed as a body tree. The body tree actually is displayed in Fig. 2(a). The tree nodes are the joints and the arcs are the rigid skeletal connections between the joints.

Since the skeleton is rigid, the human posture is uniquely specified by the angular displacements of all of the joints. Correspondingly, human motion will be uniquely represented by the angular velocities of all the joints[2]. Hence, we propose here to introduce vectorial representation of action/activity which uniquely specifies humanoid/human motion in terms of a sequence of vectors that represent the angular positions and velocities of all the joints. Since human motion may require complex sequence of varying joint angular velocities, we propose here to approximate the angular velocities by elementary motions where each segment has a constant velocity. Thus, each elementary motion may be specified by an angular joint velocity and initial and final angular displacements. A motion sequence is constructed from a temporal sequence of elementary motions. The elementary motion for the entire humanoid is actually determined by the skeletal body tree corresponding to Fig. 2(a). Each node in this skeleton tree has an associated motion sub-vector $\mathbf{J}_n$, which specifies the motion for the corresponding body joint n[3].

$$\mathbf{J}_n = [\alpha_n^i\ \phi_n^i\ \psi_n^i\ \alpha_n^f\ \phi_n^f\ \psi_n^f\ \dot{\alpha}_n\ \dot{\phi}_n\ \dot{\psi}_n]^T\ . \qquad (1)$$

$\mathbf{J}_n$ specifies the 3-D initial $(\alpha_n^i\ \phi_n^i\ \psi_n^i)$ and final $(\alpha_n^f\ \phi_n^f\ \psi_n^f)$ angular displacements and the angular velocities $(\dot{\alpha}_n\ \dot{\phi}_n\ \dot{\psi}_n)$ for the joint n along its three rotational axes. This motion data can be extracted using forward or inverse kinematics [15] [19], and corresponds to the relative motion of that joint. To specify the entire body motion, these $\mathbf{J}_n$ are arranged in a vector of motions specifying the complete set of joint motions. The motion state vector M is given by

$$\mathbf{M} = [\mathbf{J}_r, \mathbf{J}_{nb}, \cdots \mathbf{J}_{la}] \qquad (2)$$

where the subscripts like r, nb and la are the joint names as shown in Fig. 2(a). For example, la is the LeftAnkle joint.

3. THEORETICAL FOUNDATION OF OUR APPROACH

In this section, we describe the theoretical foundation to our approach in recognizing human activity using indexing. Our representation for human activity in video frames could be described as a concatenation of 18 dimensional sub-vectors $\mathbf{x}_i$ that describe the angles and angular velocities of 9 body parts[4]. Each sub-vector pertains to a video frame and thus the whole video sequence can be represented by a vector $\mathbf{Y}$ which is a concatenation of all the sub-vectors $\mathbf{x}_i$. Please note that in our representation the angles are only 2-D projections of the actual 3-D angles. Hence, our representation is limited to a given view of the activity and so our scheme is view based. However, we find that this representation is not very sensitive to changes in vantage point and the viewing direction can be changed in the range of ±30 degrees without seriously affecting the recognition rate. In the future, we plan to incorporate a method for recovery of the 3-D angles [1] that will enable us to make our recognition method view invariant.

To recognize an activity one has to compare the test video to a model activity. In other words, the test vector $\mathbf{Y}_t$ has to be compared with a set of model vectors $\{\mathbf{Y}_m; m \in [1, M]\}$ where M is the number of activity models in the database. A

[2] Given the translations and rotations of the central skeletal coordinate system as well.

[3] We model all the $\mathbf{J}_n$ with respect to the skeleton central coordinate system. Thus, all translations and rotations are entirely specified by M.

[4] The nine body parts are torso and head, upper arms and legs (thighs) and lower arms (forearm plus hand) and legs (calf plus foot).

similar problem was dealt with using Hidden Markov Models(HMM) [23] [20]. We find that the solution can be significantly simplified if we make some assumptions that will be detailed later. The problem of activity recognition can be formulated as Maximum Likelihood Sequence Estimation(MLSE). The MLSE problem is to determine the most likely sequence $\mathbf{Y_m}$ given the observations $\mathbf{Y_t}$. The Viterbi algorithm [14] provides a computational approach to solve such a problem. We assume that the random differences between the sub-vectors $\mathbf{x_t}$ and $\mathbf{x_m}$ can be described as multivariate zero mean Gaussian distribution. Assuming that these variations are conditionally independent from sample to sample, then the likelihood function for the sequence $P(\mathbf{Y_t}|\mathbf{Y_m})$ can be written as

$$P(\mathbf{Y_t}|\mathbf{Y_m}) = P(\mathbf{x_{t1}}, \mathbf{x_{t2}}, \cdots, \mathbf{x_{tk}}|\mathbf{x_{m1}}, \mathbf{x_{m2}}, \cdots, \mathbf{x_{mk}}) = \prod_{i=1}^{k} \frac{e^{[\frac{-1}{2}(\mathbf{x_{ti}}-\mathbf{x_{mi}})^T C_x^{-1}(\mathbf{x_{ti}}-\mathbf{x_{mi}})]}}{(2\pi)^{\frac{N}{2}}|C_x|^{\frac{1}{2}}} \quad (3)$$

where C_x is the covariance matrix of the distribution of the training set for $\mathbf{x_m}$, N is the dimension of the sub-vectors $\mathbf{x_m}$ or $\mathbf{x_t}$(18 in our case) and k is the number of frames in the activity sequence. Using the log-likelihood function we get

$$\log P(\mathbf{Y_t}|\mathbf{Y_m}) = \sum_{i=1}^{k}[\frac{-1}{2}(\mathbf{x_{ti}} - \mathbf{x_{mi}})^T C_x^{-1}(\mathbf{x_{ti}} - \mathbf{x_{mi}})] - kG \quad (4)$$

where G is the logarithm of the denominator in Equation 1 given by

$$G = \log[(2\pi)^{\frac{N}{2}}|C_x|^{\frac{1}{2}}] \quad (5)$$

The most likely activity sequence Ω is found by the maximum likelihood,

$$\Omega = \arg\max_m(\sum_{i=1}^{k}[\frac{-1}{2}(\mathbf{x_{ti}} - \mathbf{x_{mi}})^T C_x^{-1}(\mathbf{x_{ti}} - \mathbf{x_{mi}})]) \quad (6)$$

3.1 Foundations of the Voting approach

Finding the most likely activity can now be solved by an indexing based voting approach. In this case, for each test sub-vector $\mathbf{x_{ti}}$ we accumulate votes for all the models. In such voting, a model m will accumulate an incremental vote of

$$\frac{-1}{2}(\mathbf{x_{ti}} - \mathbf{x_{mi}})^T C_x^{-1}(\mathbf{x_{ti}} - \mathbf{x_{mi}}) - G \quad (7)$$

for each test frame i. This process is repeated by voting all the frames i in the test sequence. In our method, we even simplify this voting further by voting only on a few representative frames which are sparsely sampled from the test video sequence. As demonstrated in Section 5 four sparse samples are sufficient to achieve quite robust recognition.

3.2 Dealing with time shifts and activity speed variations

In most test sequences, we encounter the problem that the activity is not synchronized with the model activity. Usually there is a time shift between the two sequences. This time shift denoted by a, is apriori unknown and has to be found along with the activity classification. We solve this problem by combining the votes with temporal correlation.

$$\Omega = \arg\max_m(\arg\max_{a_m}(\sum_{i=1}^{k}[\frac{-1}{2}(\mathbf{x_{ti}} - \mathbf{x_{m(i-a_m)}})^T C_x^{-1}(\mathbf{x_{ti}} - \mathbf{x_{m(i-a_m)}})])) \quad (8)$$

where a_m is the time shift between the test sequence and the m^{th} model sequence of the activity. We use this method in Section 4.1 in our temporal correlation scheme.

Another problem that arises in many activities is the problem of speed variations of the activity. The same activity could be performed with different speeds and the speed can even vary during the course of the activity. Variations of speed are actually equivalent to variations in time scale. This problem is quite difficult in general since it requires complex search for the optimum votes with various time scales and time shifts.

$$\Omega = \arg\max_m(arg\max_s(\arg\max_{a_m}(\sum_{i=1}^{k}[\frac{-1}{2}(\mathbf{x_{ti}} - \mathbf{x_{ms(i-a_m)}})^T C_x^{-1}(\mathbf{x_{ti}} - \mathbf{x_{ms(i-a_m)}})]))) \quad (9)$$

where s denotes the time scale.

In Section 4.2 we propose a method which provides an optimal and robust solution to speed invariant activity recognition. Our solution is based on Sequence matching of the sparse samples. The first underlying principle in the method is that the sequence of the samples of any activity do not change with any variations of speed. This is obvious. Thus, we can reduce the search space by first searching for the optimal vote for the first test frame $\mathbf{x_{t1}}$ and then search for the next optimal vote for the second test frame $\mathbf{x_{t2}}$ only in the reduced set of model frames which occur after the matched model frame with $\mathbf{x_{t1}}$. The same process repeats with the third test frame , the fourth test frame and so on. To avoid the problem that the first test frame is matched with a model frame which occurs towards the end of the sequence we extend all the model sequences by a full additional period.

4. MULTIDIMENSIONAL INDEXING AND VOTING

In this section we describe the three different schemes that we propose for the recognition of human activity. Subsection 4.1 discusses the approach based on temporal correlation. The subsequent two subsections use only the sequence information and disregard the temporal data. The method in subsection 4.3 differs from the one in subsection 4.2 by the additional consideration of the angular velocity of the body parts.

Our activity recognition scheme is based on the fact that the range of poses of different body parts is limited and the activity patterns of the parts are largely repetitive for most of human activities. By exploiting this fact, we can reduce the redundancy in the activity database drastically. This can be achieved by decomposing the activity pattern of the whole body into a combination of activity patterns of individual body parts and storing in the same bins similar postures with different time instants. For example, a person's running activity can be regarded as a combination of the moving sequences of the arms, legs and torso. In addition, the representation is compressed by quantizing all the possible poses of body parts and representing activity patterns by sequences of discrete symbols. In this paper, these symbols are represented by multi-dimensional tuples generated from the poses of the parts. Those are later used as indices of pose hashing tables.

4.1 Indexing with Temporal Correlation

In our approach, we use a human model similar to the one used in [4] with slight variations. The human body is represented by 9 cylinders with elliptic cross-sections for the torso, upper arms and legs (thighs), and lower arms (forearms + hands) and legs (calves + feet). Furthermore, it is assumed that the Cartesian coordinates of all the major joints connecting the above mentioned parts have been obtained using a tracking procedure for body parts [4] [17] [16]. The posture of the whole body any instant is composed of the poses of the arms, legs and torso. To achieve invariance to size, the Cartesian coordinates are transformed into angles. In this method, we use 2D tuples (θ_1, θ_2) to represent the angular poses of arms and legs, where θ_1 denotes the angle between the positive x-axis and the upper arm or the thigh and θ_2 represents the angle between the positive x-axis and the forearm or the calf. For the torso, a single angle θ_3 is used for pose representation, where θ_3 represents the angle between the positive x-axis and the major axis of the torso. All the angles are measured in counter-clockwise direction. We note that the absolute spatial position of the torso in the image does not bear much activity information since it largely depends on the relative position of the imaging system.

The next step is to quantize these multi-dimensional tuples into multidimensional bins to form indices into separate hash tables. In our indexing scheme, we have five hash tables: one (h_1) for the torso, two $(h_2$ and $h_3)$ for legs and two $(h_4$ and $h_5)$ for the arms. Depending on the context, $\{h_i; i \in [1,5]\}$ are used to denote both the body part and the corresponding hash table. h_1 has one dimensional bins $b_1 = (b_{11})$ and $\{h_i : i \in [2,5]\}$ have two dimensional bins $b_i = (b_{i1}, b_{i2}), i = 2, \cdots, 5$. Each bin in the hash table contains a pair of values which denote the model number $\{m; m \in [1, M]\}$ and the time instant $\{t; t \in [0, T_m - 1]\}$ of the model activity in the database, where M is the number of activity models in the database and T_m represents the number of image frames for model m. Each hash table is updated using the angular position of the body parts obtained from each activity model. Thus, the poses of body part h_i of model m at instant t are quantized into bin b_i. The complete activity models are scattered into five hash tables (four tables for the limbs and one for the torso). In the hash table, every entry may include a set of different activity models which pertains to the same body part pose. This arrangement of the hash tables is quite efficient for storage and also enables robust recognition. Similar general principles were used in other voting schemes such as the geometric hashing [21], but our method employs several hash tables in parallel.

Our recognition scheme consists of three stages. The first stage involves voting for the individual body parts. The second stage combines the votes of the individual body parts for each test frame. The third stage obtains the final activity vote by integrating the votes of individual test frames based on the temporal information contained in the test sequence.

In the first stage, we decompose the body posture in each frame into angular poses of body parts as described above and index into the hash tables of the corresponding parts. The voting scheme for each part h_i employs M 1D arrays $V_{mk}^{h_i}(t), m \in [1, M]$, where each array corresponds to a different activity model and k is the frame number of the test activity. If we have several items in the table entry that correspond to the same pose index, most likely these items correspond to different activity models and may pertain to different time instants. Thus, the sizes of these 1D arrays should be large enough to accommodate for all time instants (i.e. time bins) of the respective activity models.

In order to tolerate slight pose variations that may occur in the same activity, it is necessary to consider also the neighboring pose bins of the indices derived from the poses of the test activity. Thus, for a given test pose, votes are accumulated from all the neighboring pose bins. The indices for pose bins are two dimensional for the hash tables of the legs and the arms, one dimension relates to the pose of upper parts (upper arms or thighs) and the other dimension denotes the pose of the lower parts (forearms or calves). Let $b_i^k = (q_1^k, q_2^k)$ denote the quantized bin of one of the limbs ($h_i, i \in [2,5]$) for a test pose in test frame k, and let $b_i' = (q_1', q_2')$ denote a neighboring bin in the corresponding hash table. We define $f(d,e)$ as a mapping function from a bin's offset d, e to the f range $[0,1]$. Here, we have chosen the mapping function to be a 2D Gaussian,

$$f(d,e) = e^{\frac{-1}{2}[(\frac{d-d_0}{\sigma_d})^2 + (\frac{e-e_0}{\sigma_e})^2]} \tag{10}$$

where σ_d, σ_e denote the scale of the Gaussian along the respective axes, (d_0, e_0) represent the center of the function. In such a case, a model m with time instant t in the entry $h_i(b_i'), i \in [2,5]$ receives a vote from the test pose according to

$$V_{mk}^{h_i}(t) \mathrel{+}= (f(|q_1^k - q_1'|, |q_2^k - q_2'|)) \tag{11}$$

where += represents incrementing the value of the left-hand side by the value of the right-hand side. $V_{mk}^{h_i}(t)$ is initialized to zero before the voting begins. This voting mechanism is illustrated in Fig. 3. For voting on the poses of the torso, we use

$$V_{mk}^{h_1}(t) \mathrel{+}= f(|q_3^k - q_3'|) \tag{12}$$

where q_3^k denotes the quantized bin of the angular pose of the torso in test frame k, q_3' denotes a neighboring bin and the mapping function f is a 1D Gaussian.

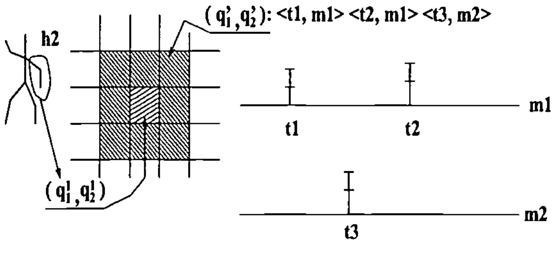

Figure 3: A voting example of the left arm. On the left, the center square (q_1^1, q_2^1) of the grid represents the bin from the pose of the left arm and the surrounding squares are neighboring bins. The upper-left bin (q_1', q_2') contains three entries from models m1 and m2. These votes are described by the bars on the right diagram. This diagram describes two 1D voting arrays for activity models m1 and m2.

In the second stage, the votes that correspond to a particular test image frame k is denoted as $V_{mk}(t)$ and are obtained by combining the votes for the torso and the votes for other body parts. The votes from the limbs and torso are combined by addition. Hence, the votes for a test image frame are given by:

$$V_{mk}(t) = V_{mk}^{h_1}(t) \ + \ \sum_{i=2}^{5} V_{mk}^{h_i}(t) \tag{13}$$

Further, if there are K number of test frames, we need to allocate one set of voting arrays ($V_{mk}^{h_i}$ and V_{mk}) for each image frame k in the test sequence and perform the same procedure as described above to gather votes from all the individual frames. The final result of the first two stages is a set of M 1D voting arrays $\{V_{mk}(t); m = 1, \cdots, M\}$ where m is the model number and k represents a test frame $k = 1, \cdots, K$.

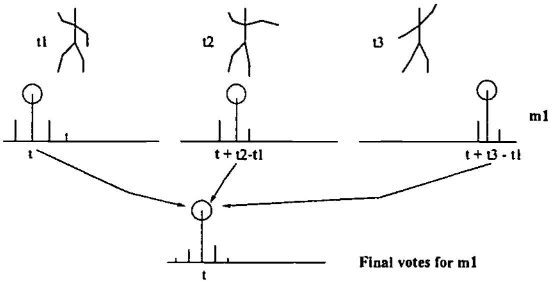

Figure 4: Using temporal correlation for integrating votes from individual test frames which provides a final voting array for one model activity.

After obtaining the votes for all the models from every distinct frame in the test sequence in the second stage, it is necessary to combine the votes from each test frame into a final vote for each activity model. This task is performed by the third stage. Let $t_1, \cdots, t_K$ represent the time instants of the frames in the test sequence, then the final vote for mth model can be obtained by the following discrete correlation

$$V_m(t) = \sum_{k=1}^{K} \sum_{\tau=-a}^{a} g(\tau)\ V_{mk}(\tau + t + t_k - t_1), t \in [0, T_m - 1] \tag{14}$$

where $g(\tau)$ is a symmetric weighting function and $[-a, a]$ is its support. An example of this stage is given in Fig. 4. The idea of temporal correlation is illustrated in Fig. 5.

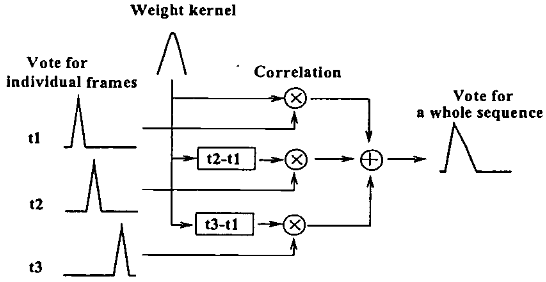

Figure 5: Combining the voting from 3 test frames in the third stage by discrete correlation.

A final scalar vote for each activity model can be obtained by

$$V_m = \max_t V_m(t) \tag{15}$$

and can be used to select one or several models which are the most similar to the test activity. The final selection also yields the exact timing of the matched activity.

4.2 Indexing with Sequence Correlation

The above temporal correlation has the shortcoming of not being able to recognize the same activity when the action is performed at different speeds. For example, fast walking may be recognized as running.

To overcome this problem, we eliminate in the second method the time component and keep only the sequence information when calculating the final vote in the third stage. When a test activity is performed at a different speed from that of the model activity, the time instants for each body pose are different. Hence the temporal correlation approach may not yield a strong response for the recognized activity which may lead to false alarms in the recognition system. In the second approach, we attempt to create speed-invariant activity recognition by eliminating the temporal information and replacing it by the sequence of the activity.

This method uses the same indexing scheme discussed above. The difference arises in the third stage. Here, we extend the vote table obtained in the second stage so that it is equivalent to the activity with two cycles instead of one. In the third stage, the final vote for the mth model can be obtained by the following equation

$$V_m = \sum_{k=1}^{K} V_{m,k}(L_k) \tag{16}$$

In the above equation the following conditions have to be satisfied: $L_i < L_j;\ i < j$ and $V_{m,k}(L_k)$ is the maximum vote for the activity m.

4.3 Sequencing with Angular Velocity

We develop this method to improve the discrimination between activities as compared to the method in 2.2. In this method, the multidimensional hash table is extended to 4 dimensions instead of two for the limbs and 2 dimensions instead of one for the torso. The additional two dimensions for the limbs are used to represent their angular velocities. We use 4D tuples $(\theta_1, \theta_2, \dot{\theta}_1 \dot{\theta}_2)$, where θ_1 and θ_2 are same as for method one, $\dot{\theta}_1$ denotes the angular velocity of the upper arm or thigh and $\dot{\theta}_2$ denotes the angular velocity of the forearm or the calf. For the torso, the 2D tuples are, $(\theta_3, \dot{\theta}_3)$, where θ_3 is the same as method one and $\dot{\theta}_3$ is the angular velocity of the torso. The angular velocities are calculated as the difference of the angular positions of two successive frames. For the last frame, the angular velocity of the last but one frame is retained. The angular velocities are then quantized into 5 bins.

In this new indexing scheme, the four hash tables for the limbs are now indexed using four dimensions and the hash table for the torso is indexed using two dimensions. The bins in each table contain a pair of model number and frame number as before.

The first stage of voting is similar as in the previous methods where the body posture of each frame is decomposed into the poses of the five body parts and indexed in to the hash tables of the corresponding parts. In order to tolerate slight pose variations that may occur in the same activity, it is necessary to consider also the neighboring pose bins of the indices derived from the poses of the test activity. Let $b_i^k = (q_1^k, q_2^k, q_3^k, q_4^k)$ denote the quantized bin of one of the limbs ($h_i, i \in [2,5]$) for a test pose in test frame k, and let $b_i' = (q_1', q_2', q_3', q_4')$ denote a neighboring bin in the corresponding hash table. We define $f(b,c,d,e)$ as a mapping function from a bin's offset d, e to the f range $[0,1]$. Here, we choose this mapping function to be a 4D Gaussian,

$$f(b,c,d,e) = e^{\frac{-1}{2}[(\frac{b-b_0}{\sigma_b})^2 + (\frac{c-c_0}{\sigma_c})^2 + (\frac{d-d_0}{\sigma_d})^2 + (\frac{e-e_0}{\sigma_e})^2]} \qquad (17)$$

where $\sigma_b, \sigma_c, \sigma d, \sigma e$ denote the scale of the Gaussian along the respective axes, (b_0, c_0, d_0, e_0) represent the center of the function. In such a case, a model m with time instant t in the entry $h_i(b_i'), i \in [2,5]$ receives a vote from the test pose according to

$$V_{mk}^{h_i}(t) += \alpha(f(|q_1^k - q_1'|, |q_2^k - q_2'|, |q_3^k - q_3'|, |q_4^k - q_4'|)) \qquad (18)$$

where += represents incrementing the value of the left-hand side by the value of the right-hand side. $\alpha = 1, if\ |q_3^k - q_3'| = 0\ and\ |q_4^k - q_4'| = 0\ \alpha = 0.5,\ if\ |q_3^k - q_3'| = 1\ and\ |q_4^k - q_4'| = 0$ or $if\ |q_3^k - q_3'| = 0\ and\ |q_4^k - q_4'| = 1$

$V_{mk}^{h_i}(t)$ is initialized to zero before the voting begins. This voting mechanism is illustrated in Fig. 3. For voting on the poses of the torso, we use

$$V_{mk}^{h_1}(t) += \alpha\ f(|q_5^k - q_5'|, |q_6^k - q_6'|) \qquad (19)$$

where q_5^k denotes the quantized bin of the angular pose of the torso in test frame k, q_6^k denotes the quantized bin of the angular velocity of the torso in test frame k, q_5' , q_6' denote a neighboring bin and the mapping function f is a 2D Gaussian. $\alpha = 1, if |q_6^k - q_6'| = 0\ \alpha = 0, if |q_6^k - q_6'| \neq 0$

The second and third stages are identical to the corresponding stages in method 2.

5. EXPERIMENTAL RESULTS

We apply our three methods to a database of eight different human activities. These activities are jumping, kneeling, picking, putting, running, sitting, standing and walking. A total of 26 activity sequences are stored in the database. For each activity we have three or four different sequences performed by different persons. Fig. 6 and Fig. 7 shows a few sample frames of such activity sequences.

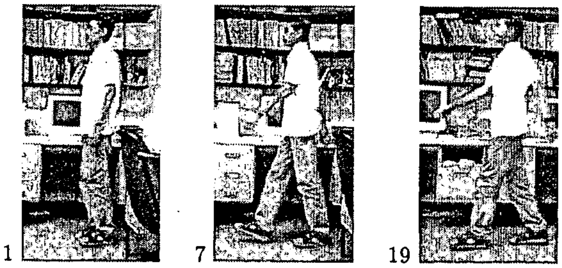

Figure 6: 3 frames of the 'walking' activity. The numbers on the lower-left corners indicate frame numbers.

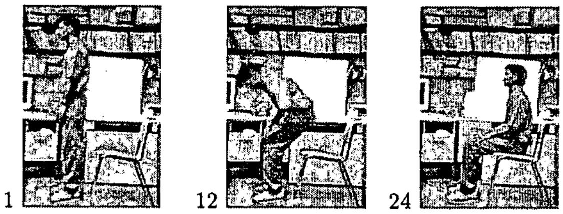

Figure 7: 3 frames of the 'sitting' activity. The numbers on the lower-left corners indicate frame numbers.

The test sequences are generated by taking three or four frames from the model activity sequences sampled at uneven time intervals and adding random perturbations to the positions of the body parts in the frames. We generate four test sequences for each kind of activity. These test sequences are matched against all the activity sequences in the database except the one from which the test sequence is extracted.

Table 1 displays the average votes and the standard deviation for each possible activity pairs for the first method.

This method may discriminate between activities performed at different speeds because the time instants for each body pose are different. This may result in an increase in the false alarm rate. We improve this method by eliminating the time factor and considering only the sequence information and hence make it invariant to speed. The results for this sequence based voting are shown in Table 2. We observe that this method helps in recognizing activities from the database that are performed at different speeds.

In order to discriminate between the activities with more accuracy we introduce the angular velocity of the body parts. The results of this method are shown in Table 3. It is observed that the discrimination between the activities improves. The drawback of this method is that the recognition is slower than the other two methods due to the increased size of the hash tables.

	Jump	Kneel	Pick	Put	Run	Sit	Stand	Walk
Jump	**12.35** (1.26)	8.24 (1.30)	6.51 (0.89)	6.06 (0.74)	6.00 (1.00)	5.85 (1.19)	4.60 (1.07)	7.79 (1.00)
Kneel	9.67 (0.97)	**13.3** (0.60)	7.32 (1.85)	5.91 (0.35)	5.78 (0.74)	7.16 (1.00)	6.02 (1.44)	8.72 (0.70)
Pick	3.26 (0.89)	4.90 (0.58)	**9.57** (0.24)	6.23 (0.38)	3.64 (0.92)	6.14 (0.67)	8.19 (0.40)	4.43 (0.46)
Put	5.57 (0.64)	5.77 (0.62)	8.15 (1.50)	**13.0** (1.04)	6.06 (1.25)	7.56 (1.11)	6.27 (0.90)	6.13 (0.81)
Run	5.58 (0.74)	5.52 (0.91)	5.98 (1.12)	6.44 (1.73)	**7.46** (1.28)	4.97 (0.98)	5.02 (1.42)	6.70 (1.37)
Sit	3.97 (1.32)	5.14 (1.35)	7.58 (1.85)	7.82 (1.22)	4.76 (1.68)	**12.8** (0.86)	9.11 (1.46)	5.69 (1.55)
Stand	1.12 (0.47)	1.66 (0.48)	3.56 (0.93)	5.87 (0.94)	3.67 (0.98)	7.48 (1.06)	**12.3** (0.26)	2.19 (0.59)
Walk	7.64 (0.56)	7.13 (0.67)	6.31 (1.56)	5.09 (1.12)	6.60 (1.55)	6.30 (1.33)	6.00 (1.18)	**9.60** (1.96)

Table 1: Average votes (Standard deviation) of activity sequences for the temporal correlation based voting. The rows correspond to test activity while the columns correspond to the model activities. The best score in each row is in boldface numerals. The method yields correct recognition since the scores along the diagonal are the highest in each row.

	Jump	Kneel	Pick	Put	Run	Sit	Stand	Walk
Jump	**15.37** (0.71)	7.00 (0.88)	6.00 (0.42)	6.38 (0.24)	5.38 (1.12)	6.54 (0.42)	3.90 (0.84)	8.10 (1.41)
Kneel	10.9 (2.42)	**16.0** (0.81)	7.60 (2.47)	6.91 (0.41)	5.80 (2.07)	9.60 (0.97)	5.63 (1.63)	9.58 (1.73)
Pick	3.45 (2.12)	4.77 (2.10)	**11.7** (0.57)	8.00 (2.30)	5.14 (2.30)	6.00 (2.10)	8.35 (1.15)	5.31 (2.31)
Put	5.30 (1.34)	5.57 (1.57)	10.3 (2.24)	**15.0** (1.57)	6.19 (2.21)	8.90 (2.19)	6.80 (1.73)	7.75 (1.31)
Run	4.06 (0.71)	4.25 (0.73)	3.96 (1.48)	5.20 (1.68)	**7.21** (1.89)	4.60 (1.26)	4.85 (1.62)	5.84 (1.41)
Sit	4.41 (1.00)	5.01 (1.16)	7.97 (1.99)	8.04 (2.14)	6.22 (2.45)	**15.0** (0.71)	10.8 (1.61)	6.28 (1.70)
Stand	2.42 (3.10)	3.08 (2.24)	6.16 (3.16)	6.75 (2.50)	5.51 (2.00)	10.1 (1.41)	**15.1** (0.63)	4.15 (4.00)
Walk	7.60 (2.17)	7.66 (1.14)	4.80 (0.88)	5.82 (0.31)	7.27 (2.00)	7.47 (1.40)	5.00 (1.67)	**12.3** (1.18)

Table 2: Average votes (Standard deviation) of activity sequences for the sequence based voting. The rows correspond to test activity while the columns correspond to the model activities. The best score in each row is in boldface numerals. The method yields correct recognition since the scores along the diagonal are the highest in each row.

	Jump	Kneel	Pick	Put	Run	Sit	Stand	Walk
Jump	**12.31** (1.21)	3.91 (0.51)	1.97 (0.54)	2.00 (0.45)	2.18 (0.60)	2.00 (0.55)	1.20 (0.95)	3.55 (0.78)
Kneel	4.90 (2.16)	**9.99** (2.18)	3.20 (0.88)	2.77 (0.63)	2.20 (0.77)	2.18 (0.71)	2.40 (1.58)	3.80 (0.92)
Pick	0.67 (0.72)	2.00 (0.60)	**8.00** (0.71)	2.40 (0.85)	1.97 (1.00)	1.90 (0.86)	3.80 (0.96)	1.36 (0.95)
Put	1.95 (0.73)	2.58 (0.71)	3.10 (0.51)	**8.37** (2.95)	1.58 (1.13)	4.71 (1.00)	2.50 (0.81)	1.74 (0.90)
Run	2.00 (0.75)	2.25 (0.83)	1.40 (0.37)	1.70 (0.32)	**3.23** (1.28)	1.40 (0.28)	1.50 (0.39)	2.90 (0.94)
Sit	1.36 (0.47)	1.73 (0.45)	3.00 (0.67)	4.20 (0.94)	0.90 (0.95)	**8.60** (0.83)	3.40 (1.48)	1.60 (0.63)
Stand	0.00 (0.00)	0.55 (1.04)	2.34 (1.83)	1.86 (1.00)	1.23 (1.17)	3.50 (0.67)	**9.90** (0.37)	0.63 (1.11)
Walk	3.40 (0.99)	3.18 (1.07)	1.97 (0.87)	1.60 (1.05)	2.61 (1.09)	1.50 (0.44)	1.16 (0.69)	**5.75** (2.20)

Table 3: Average votes (Standard deviation) of activity sequences for the sequence with angular velocity based voting. The rows correspond to test activity while the columns correspond to the model activities. The best score in each row is in boldface numerals. The method yields correct recognition since the scores along the diagonal are the highest in each row.

6. CONCLUSIONS

In this paper, we propose and evaluate a representation for human activity/action that is based on sequences of angular poses/velocities of the human skeletal joints. The evaluation is implemented by developing a multidimensional indexing scheme for activity retrieval and storage. For this purpose, we develop three different approaches to human activity recognition/retrieval which are based on this representation. The sequence based voting approach in the second and third methods, is introduced since the temporal correlation approach in the first method , is not invariant to speed and incorrectly recognizes running activity as walking in one case. The second method solves this problem, but at the expense of incorrectly recognizing the stand activity as sitting in one case. This happens because we take two cycles during the voting process. These two activities differ only in the sequence with which they occur. Hence the misclassification. When we introduce the angular velocity in the third method this misclassification is no longer present and in fact it gives better discrimination between the activities as it is expected, due to the increased dimensionality. To evaluate the effectiveness of the methods quantitatively we define the Average Discrimination Ratio (ADR) as the average of the ratios of the first maximum vote to the second maximum vote for each activity. The ADR for the three methods are 1.38, 1.49 and 2.15 respectively. This shows that the third method has the best discrimination power.

In summation, we propose here a representation for human action/activity which can describe accurately any complex human activity/action and develop a robust method for activity recognition/retrieval. The indexing approach also provides an efficient storage/retrieval of all the activities in a small set of hash tables.

7. ACKNOWLEDGMENTS

This work was supported by the National Science Foundation (NSF) Grants No. IIS-9711925, IIS-9876904 and IIS-9979774.

8. REFERENCES

[1] C. Barron and I. A. Kakadiaris. Estimation anthropometry and pose from a single image. *Proc. Conf. Computer Vision and Pattern Recognition*, pages 669–676, June 2000.

[2] C. Bregler and J. Mallik. Tracking people with twists and exponential maps. *Proc. IEEE 1998 Int'l Conf. Computer Vision and Pattern Recognition (CVPR'98)*, pages 8–15, June 1998.

[3] S. K. Chang, Q. Y. Shi, and C. W. Yan. Iconic indexing by 2-d strings. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(3):413–427, July 1987.

[4] L. Concalves, E. Bernardo, E. Ursella, and P. Perona. Monocular tracking of human arm in 3d. *Proc. 1995 International Conference on Computer Vision*, 1995.

[5] A. Del-Bimbo, E. Vicario, and D. Zingoni. Symbolic description and visual querying of image sequences using spatiotemporal logic. *IEEE Transactions on Knowledge and Data Engineering*, 7(4):609–622, August 1995.

[6] D. Gavrila and L. Davis. Towards 3-d model-based tracking and recognition of human movements. *Proc. of the 1995 Int. Workshop on Automatic Face and Gesture Recognition*, 1995.

[7] A. H. Guest. Dance notation: The process of recording movement on paper. *London:Dance Books*, 1984.

[8] A. H. Guest. Chore-graphics: A comparison of dance notation systems from the fifteenth century to the present. *Systems from the Fifteenth Century to the Present, Gordon and Breach Science Publishers S. A.*, 1989.

[9] D. Herbison-Evans. Dance, video, notation and computers. *Leonardo*, 1988.

[10] D. Hogg. A program to see a walking person. *Image and Vision Computing*, 5(20), 1983.

[11] S. JU, M. Black, and Y. Yacoob. Cardboard people: A parameterized model of articulated motion. *2nd Int. Conf. on Automatic Face and Gesture Recognition*, pages 38–44, 1996.

[12] E. Jungert. The observer's point of view, an extension of sympolic projections. *Prof.In. Conf. of Theories and Methods of Spatio-Teporal Reasoning in Geopraphic Space*, pages 179–195, September 1992.

[13] I. A. Kakadiaris and D. Metaxas. Model-based estimation of 3d human motion with occlusion based on active multi-viewpoint selection. *Proc. 1996 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'96)*, 1996.

[14] T. K. Moon and W. C. Stirling. *Mathematical Methods and Algorithms for Signal Processing.* Prentice Hall Inc., 2000.

[15] R. M. Murray, Z. Li, and S. S. Sastry. *A Mathematical Approach to Robotic Manipulation.* CRC Press Inc., 1994.

[16] T. M. Naoyuki Sawasaki and T. Uchiyama. Design and implementation of high-speed visual tracking system for real-time motion analysis. *Proc. of ICPR 1996*, pages 478–484, 1996.

[17] J. Regh and T. Kanade. Model-based tracking of self-occluding articulated objects. *Proc. 1995 International Conference on Computer Vision*, 1995.

[18] K. Rohr. Incremental recognition of pedestrians from image sequences. *Proc. 1995 Comp. Soc. Conference on Computer Vision and Pattern Recognition*, pages 8–13, June 1993.

[19] R. J. Schilling. *Fundamentals of Robotic Analysis & Control.* Prentice Hall Inc., 1990.

[20] J. Schlenzig, E. Hunter, and R. Jain. Vision based hand gesture interpretation using recursive estimation. *Proceedings of the 28th Asilmoar Conference on Signals, Systems and Computers*, 1994.

[21] J. Schwartz, Y. Lamdan, and H. Wolfson. Geometric hashing : A general and efficient model-based recognition scheme. *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pages 335–344, 1988.

[22] A. P. Sistla and O. Wolfson. Temporal triggers in active database systems. *IEEE Transactions on Knowledge and Data Engineering*, 7(3), June 1995.

[23] T. Starner and A. Pentland. Real time american sign language recognition from video using hidden markov models. *Proceedings of the International Symposium on Computer Vision*, 1996.

[24] A. D. Wilson, A. F. Bobick, and J. Cassell. Temporal classification of natural gesture and application to video coding. *IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recogni tion*, pages 948–854, June 1997.